

カードゲームを対象とした セマンティック データの使用方法に関する研究

00YA, Fumihiro / 大家, 史裕

(出版者 / Publisher)

法政大学大学院理工学・工学研究科

(雑誌名 / Journal or Publication Title)

法政大学大学院紀要. 理工学・工学研究科編 / 法政大学大学院紀要. 理工学・工学研究科編

(巻 / Volume)

57

(開始ページ / Start Page)

1

(終了ページ / End Page)

5

(発行年 / Year)

2016-03-24

(URL)

<https://doi.org/10.15002/00013271>

カードゲームを対象とした セマンティックデータの利用方法に関する研究

RESEARCH ON THE UTILIZATION OF SEMANTIC DATA ON CARDGAME

大家史裕

FUMIHIRO OOYA

指導教員 藤井章博

法政大学大学院理工学研究科応用情報工学専攻修士課程

In recent years, variety of games exist either in analog or digital format. They become more than amusement, even a professional job with large prizes. In this paper, a trading card game is treated. In order to obtain beneficial information about cards in which there are so many types and numbers in each type, the machine processing method is introduced. The method is based on semantic web technology over information about cards' data extracted web database.

Key Words : *Semantic web, natural language processing, inference*

1. はじめに

近年、世界的にデジタル・アナログゲームは共に単なる娯楽の枠を超えて賞金制の大会が行われている。大会で勝利を重ねることは賞金や他の大会への招待、スポンサーの獲得に繋がり、個人に対して莫大な利益になりうる。日本でも将棋や囲碁や麻雀などがプロ制度のある競技として認知されているが、昨今ではその他のアナログゲームやデジタルゲームにおけるプロチームの発足や企業によるスポンサーが行われ始めている。

この分野では、田中哲郎氏の「どうぶつ将棋の完全解析」[1]、古井敬大氏の「相手の抽象化による多人数ポーカーの戦略の決定」[2]といったゲームの解析からコンピュータによるゲームの自動化といった幅広い研究がなされている。

そのアナログゲームの一つとしてトレーディングカードゲーム (TCG) というものがある。このゲームは囲碁や将棋と異なり、互いに使用する駒 (カード) が非対称である。したがって、定石や戦術といったカードの使い方 (プレイングスキル) だけではなく、優れたデッキ (カードの束) の構築をすることも同時に求められるため、カードやカード間の関係に対する深い知識が求められている。

本研究ではセマンティックウェブの技術を用いて、セマンティクスを付与したカードゲームに使用するカードのデータベースの作成を行う。本研究では米 Wizard of the Coast 社によって制作・販売がされている Magic: The Gathering (MTG) というカードゲームを対象とする。このゲームに使用されるカードは1万5千種類を超え、こ

れらの組み合わせは膨大な量になるためカードやカード間の有効な組み合わせを見落とす可能性がある。また1万5千種類以上あり、新規のカードが出ると一度に200枚前後追加されるため、手動でカードをRDFにする手間は相当なものだと考えられる。そこで、まずカードが持つ名前や能力といった特性をメタデータとして付与したRDFを自動で生成するプログラムを作成する。そして予め作成しておいたMTG用のオントロジーを利用して推論を行うことで、大量のカードの分類やカード間の関係を機械処理し、プレイヤーに有益になる情報を得ることを目的とする。

2. 要素技術

カードデータのRDF化や推論を行うためにセマンティックWebの技術を用いた。また、文章をRDFにする際に、機械処理ができる形にするために自然言語処理を導入した。

(1) セマンティック

セマンティックは一般的には意味論を表す言葉であるが、ITの分野では文書などのデータが持つ意味をメタデータの付与などによって解釈をさせることでコンピュータに機械処理を可能にする技術を表す言葉として用いられる。これを実現させる具体的な方法について以下で述べる。

a) RDF(Resource Description Framework)

RDFはウェブ上にあるリソースを記述するための統

一された枠組みであり、セマンティック Web を実現するための構成要素の一つである。RDF はメタデータについて記述することを目的としており、トリプルと呼ばれる主語、述語、目的語の三つ組みの要素で記述される。主語が記述対象のリソース、目的語が主語と関係のあるリソースまたはリテラル、述語が主語と目的語のリソースの関係をそれぞれ示している。例として「ザッハトルテはフランツ・ザッハによって生み出された。」という文を RDF のトリプルとして表現し、グラフ理論に則って表現したグラフを図 1 として示す。

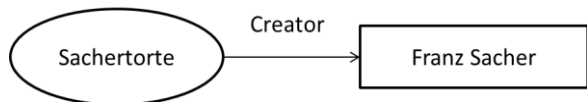


図 1 RDF の例

b) オントロジー

オントロジーとは情報科学において、ある領域の概念と概念との関係を体系的にまとめて記述したものである。オントロジーを用いることで、Web 空間上にあるテキストなどのリソースの意味やリソース間の関係を機械処理することが可能になる。

オントロジーは主に、ウェブオントロジー言語 OWL(Web Ontology Language)という RDF の語彙拡張を用いて記述される。本研究においては、OWL を扱うツールとして広く利用されている Protégé[3]を用いた。

(2) 自然言語処理

自然言語処理とは人間が普段使用している言語（自然言語）を機械処理するための技術のことを指し示す。自然言語処理にはいくつかの基礎技術が含まれるが、ここでは本研究に関係するものを抜粋して説明を行う。

a) 形態素解析

形態素解析とは自然言語の文章を、意味を持つ最小限の単位（トークン）に区切り、あらかじめ用意した辞書を用いて品詞や内容の識別を行い、トークンに品詞タグを付与する方法である。形態素解析によって得られる結果の例は図 2 の中の黒枠内である。

b) チャンキング

チャンキングとは形態素解析によって品詞タグをつけられた文章から、名詞句や動詞句といった、ある程度機能の固まった単語の列（チャンク）を取り出す方法である。チャンキングによって得られる結果の例は図 2 の中のオレンジの枠内である。この例では名詞句（NP）をチャンキングしている。

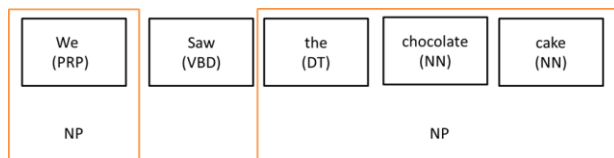


図 2 トークンとチャンクの分割とラベル付け

3. カードデータの RDF 化

前章で説明した要素技術を用いて作成するカードデータを RDF 化する手順をまとめたものを図 3 に示す。

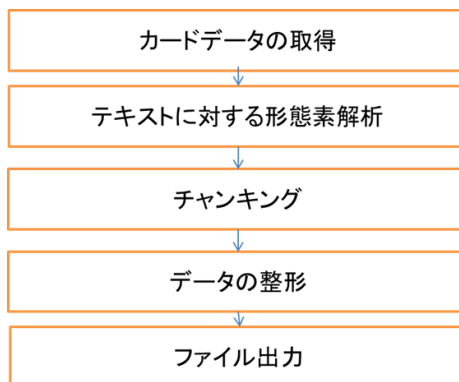


図 3 RDF 化の手順

カードデータの取得は、Web 上のページよりスクレイピングを行うことでカードのデータを入手する。

テキストに対する形態素解析は python の NLTK (Natural Language Toolkit) を用いて、単語をトークンにし、その後トークンに品詞（名詞、動詞）を表すタグ付けを行う。

チャンキングでは NLTK を用いて、形態素解析を行うことでタグ付けを行った単語列を利用してチャンキングを行う。

データの整形では RDF/XML の形に整形をし、ファイル出力で rdf ファイルとして出力する。

(1) 対象となるカードの説明

MTG において使用されるカードの中にはカード名や能力に関しての文章などのカードの特性が含まれている。以下の図 4 を使用して説明を行う。



図 4 カードの例

- 1)カード名：カードの名前。同名で特性が異なるカードは存在しない。
- 2)mana・コスト：このカードを使うために必要なコスト（mana）の数が記載されている。この場合は赤2つと不特定な4つの計6つのmanaが必要であるとなる。
- 3)タイプ：このカードの種類に関する情報が記載される。このカードの場合はクリーチャーの中のドラゴンのカードであるとなる。
- 4)エキスパンション・シンボル：カードが収録されている商品名（エキスパンション）についての記載。また色によって希少度（レアリティ）も表現されている。
- 5)能力：カードが行う特殊な行動について記載されている。
- 6)フレーバー・テキスト：ルールに関係しない、MTGのストーリーなどについてのテキスト。
- 7)パワー/タフネス:そのカードの攻撃力（パワー）と体力（タフネス）という特性について記述される。クリーチャーにのみ存在する。

(2) カードデータの入手元

機械処理を行い、有益な情報を得るためには誤ったカードの特性が記述されていることは重大な障害に成り得る。公式カードデータベースである Gatherer[4]にはルール上の間違いなどを訂正した最新のカード情報が載っている。よって、本研究ではカードデータは Gatherer のページを利用する。

(3) カードデータの RDF 化

カードを RDF 化するために Gatherer に記載されているカードの特性を取得する。カードの特性の取得を行うために、カードについて記述されているページの URL とカードの特性について記述されている部分の DIV タグの ID を利用してスクレイピングを行う。

このようにしてスクレイピングしたカード特性を RDF の形式に整形し、出力することでカードデータの RDF 化を実現する。この際に URI の略文字列である接頭辞を cardterms とし、使用するプロパティの一覧を表 1 に、このプロパティを使用した RDF のグラフを図 5 に示す。

表 1 RDF に使用したプロパティ一覧

cardName	カード名
ManaCost	mana・コスト
ConvertedManaCost	点数で見たmana・コスト
cardType	カード・タイプ
SubType	サブ・タイプ
cardText	能力
FlavorText	フレーバーテキスト
Power	パワー
Toughness	タフネス
Loyalty	忠誠度
Rarity	希少度
Expansion	エキスパンション
CollectorNumber	コレクター番号



図 5 作成した RDF のグラフ

カードが持つ特性がリテラルとして付与されている。これによって、カードデータの基本的な RDF 化は達成できた。

(4) 自然言語処理の導入

対象となるカードは能力を持っており、これは文章として表現されている。「Shivan Dragon」は「Flying」と「{R}: Shivan Dragon gets +1/+0 until end of turn.」（{R}は を表している。）という2つの能力を持っている。人間がこの能力を読んで理解することは容易いが、機械がこの能力を理解するためには自然言語処理を施す必要がある。自然言語処理はオライリー・ジャパン社の「入門自然言語処理」[5]の本の手法を参考にした。

a) テキストに対する形態素解析

まずスクレイピングによって取得したテキストに形態素解析を施す。形態素解析では、単語をトークン（言語学上の識別可能な単位）にし、その後トークンに品詞（名詞、動詞）を表すタグ付けを行う。

具体的なトークン化の例は「It' s」や「don' t」といった省略形の分割（この場合は「It ' s」「do n' t」となる。）や文末のピリオドを切り離すことが挙げられる。

トークン化をした後、品詞タグ付けを行う。単語列に品詞タグ付けを行うものを品詞タガー（または POS タガー）と呼ぶ。本研究では表 2 に示す品詞タグセットを使用した。

表 2 使用した品詞タグセット

DT	冠詞
CD	数字
NN	名詞 (単数)
NNS	名詞 (複数)
NNP	固有名詞
PRP	人称代名詞
PRP\$	所有代名詞
POS	s
RB	副詞
JJ	形容詞
MD	法
VB	動詞 (原形)
VBP	動詞 (三人称単数以外の現在形)
VBZ	動詞 (三人称単数の現在形)
VBD	動詞 (過去形)
VBN	動詞 (過去分詞)
VBG	動詞 (動名詞または現在分詞)
WP	関係代名詞
TO	前置詞 to
IN	前置詞または従属接続詞
CC	等位接続詞
SYM	記号
"	引用符
(左側の括弧
)	右側の括弧
.	コンマ
.	文の終わりの句読点
:	文の途中の句読点
IN-IF	前置詞IF
IN-TR	誘発型能力の条件
NNP-KY	キーワード能力
NNP-AW	能力語
NNP-TL	カード名

実際にタグ付けを行った例として、「Shivan Dragon」のテキストをこの手法で形態素解析した結果を図6に示す。

```
[(['', '('), ('R', 'SYM'), ('', ')'), (':', ':'), ('Shivan', 'NNP-TL'), ('Dragon', 'NNP-TL'), ('gets', 'VBZ'), ('+1/+0', 'CD'), ('until', 'IN-CE'), ('end', 'NN'), ('of', 'OF'), ('turn', 'NN'), (',', ',')]]
```

図6 形態素解析の結果

b) チャンキングの導入

この項ではチャンキングを導入する。チャンキングを行うものをチャンカーと呼ぶ。MTGのテキストはある程度定型化されているため、本研究ではチャンキングに正規表現チャンカーを用いる。正規表現チャンカーは正規表現で書かれたルールを上からチェックしていき、チャンクを順々に更新していく。そして全てのルールがチェックされると、チャンク構造が返される。NPチャンク(名詞句)は冠詞、形容詞、名詞からなる単語列だというルールを正規表現で表すと以下ようになる。

NP:{<DT|JJ|NN.*>+}

このようにルールの記述を行い、チャンクの生成を行う。表3に今回定義した主なチャンクについてまとめた。

表3 使用したチャンクの一覧

NP	名詞句
VP	動詞句
PP	前置詞句
CLAUSE	名詞句と動詞句の連なり
NPP	名詞と前置詞の連なり
MN	マナシンボル
KY	キーワード能力
AW	能力語
TE	誘発条件
CO	if節
AC	起動能力のコスト

実際の例として「Shivan Dragon」のテキストをチャンキングしたものを図7に示す。

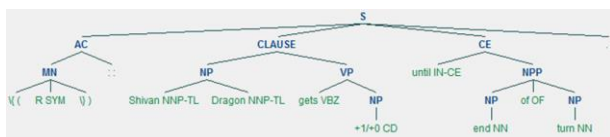


図7 チャンキングの結果

チャンクは木構造になっており、チャンクが入れ子になることもある。これまでの処理によって、カードのテキスト部の解析を達成できた。

(5) テキスト部分の RDF 化

チャンキングしたテキストデータを基に、テキスト部分のRDF化を行う。チャンキングされたデータは図7のような木構造になっている。よってこの木を走査し、チャンクの種類に応じて処理を行うことでテキスト部分のRDF化が達成される。

テキスト部分のRDFは新たに cardterms:hasAbility プロパティを導入し、カードの持っている能力を表す「[カード名]-[番号]」という形式のリソースを導入する。「Shivan Dragon」ならば能力を2つ持っているので「Shivan_Dragon-1」というリソースと「Shivan_Dragon-2」というリソースが cardterms:hasAbility プロパティによって結ばれる。その後、表4に示すプロパティ群を使用して、各カードの持っている能力を表現する。

表.4 テキスト部分に使用したRDFの例

AbilityName	能力の動作
AbilityObject	能力の主格
AbilitySubject	能力の対象
AbilityCondition	能力の条件
AbilityCost	能力のコスト
AbilityWord	能力語
ManaCost	マナ・コスト
KeywordAbility	キーワード能力
When/Whenever	誘発条件
Until	能力の期限

この手法で「Shivan Dragon」をRDF化したものを図8に示す。

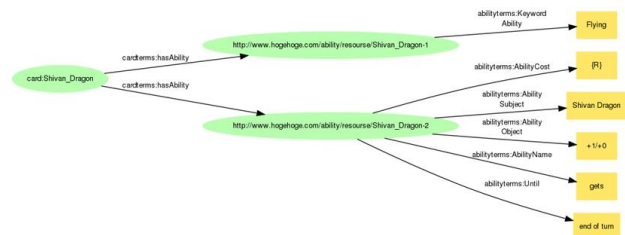


図.8 テキスト部分のRDF化の結果

これによって、カードのテキスト部分のRDF化は達成された。カードの番号と枚数を入力することで、これら一連の流れを行ってRDFに落とし込むプログラムを作成したことにより、カードデータのRDF化は達成した。

4. オントロジーの作成

作成したカードデータのRDFをより活用するためには、カードの持つ特性や特性間に関する記述が必要であり、これをオントロジーによって補完することでカードの機械処理が達成されると考える。本研究ではオン

トロジー構築ツール「Protégé」を用いて作成を行った。

(1) オントロジーの構造

作成を行ったオントロジーは MTG クラスを最上位クラスとし、その下位クラスとして Cards クラスと Glossary クラスを定義した。Cards クラスにはカードの分類に関するクラスが定義されており、Glossary クラスには MTG に関する個体の関係が記述されている。詳細については次項以降で述べる。

(2) Glossary クラス

このクラスにはカードの持ち得る特性やルール、今まで発売されたエキスパンションといった MTG に関する概念が階層化されており、実際の事柄（事実）については個体(Individual)として記述されている。

実際の例として図4で説明したカードの特性の中のタイプ行をあげる。タイプ行は「特殊タイプ」、「カード・タイプ」、「サブ・タイプ」の三種類からなり、サブ・タイプはさらに「アーティファクト」、「クリーチャー」、「エンチャント」、「土地」、「次元」、「プレインズウォーカー」、「呪文」からなる。そしてタイプ行における事実をインスタンスである個体によって記述する。「Shivan Dragon」を例にとると「クリーチャー」というカード・タイプや「ドラゴン」というサブ・タイプが個体にあたる。この概念を、クラス階層と個体の記述を用いて定義すると図9のようになる。

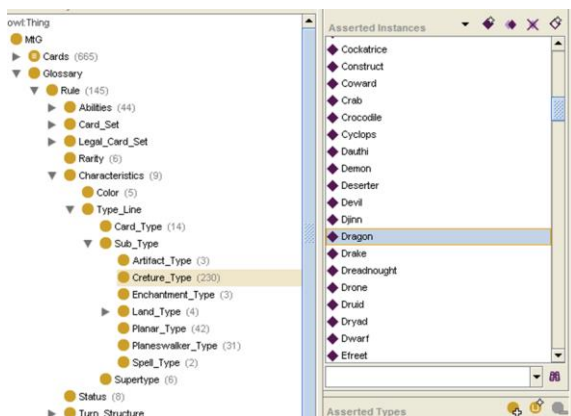


図9 クラス階層と個体

クリーチャータイプには「ドラゴン」という個体を含め、230種類が存在することを表現されている。このようにして MTG に関する事柄をオントロジー上で表現する。

(3) Cards クラス

このクラスではカードの持つ特性によってのカードの分類を、出現回数や値の範囲といったプロパティの制約条件とその論理的な組み合わせを用いた定義できる。を利用して行う。個体を用いた推論を行うために、前項で作成した RDF に、リテラルでの特性の記述の他にオント

ロジー上の個体による特性を用いた表現が必要となるため、新たに表5のプロパティを導入してトリプルを形成する。このプロパティの接頭辞は cardobjects とする。

表5 Owl のためのプロパティ一覧

プロパティ名	意味
hascardType	カードタイプ
hasSubType	サブ・タイプ
hasExpansion	エキスパンション
hasRarity	稀少度
AbilityTag	能力に関する語

AbilityTag プロパティは能力を分類に利用するために新たに導入するもので、値はその能力の動作となる動詞（引く、破壊する、与えるなど）やキーワード能力や能力語（飛行、防衛、強襲など）、能力の対象（カード、クリーチャー、対戦相手など）を対象とする。

例として「ライフゲイン（回復すること）を能力として持つカードは Gain と Life という個体をインスタンスとして持つ」という概念を Protégé 上で記述すると図10のようになる。

```
cardobjects:AbilityTag has Life
cardobjects:AbilityTag has Gain
```

図10 集合演算を用いたクラス公理

これによって、「Life」と「Gain」を持つ個体はこのクラスのインスタンスであるという記述ができた。このクラスを用いることでライフゲインを能力として持つカードを推論によって分類できる。このようにしてカードの分類についてのクラスを記述した。

5. 提案手法

3章で Web 上のページである Gatherer からカードデータをスクレイピングし、自然言語処理を用いて RDF 化をした。4章で MTG に関するオントロジーを作成し、MTG に関する事実をクラス階層と個体によって表現し、クラス公理を用いてカードの分類についてのクラスを記述した。これにより RDF とオントロジーを用いてカードに関するセマンティックを導入したデータベースが準備できた。

本章では、カードゲームのプレイヤーを支援する方法について検討する。膨大なカードの中から、有効なカードの情報を提示することはプレイヤーにとって重要である。この目的のためにセマンティックを利用したカードの分類を、個体の推論を用いて行う。

今回の推論ではエキスパンションの「タルキール霸王譚 (Khans of Tarkir)」、「運命再編(Fate Reforged)」、「タルキール龍紀伝 (Dragons of Tarkir)」に収録されている 665 種類のカードを用いて行う。

まず、これらのカードデータは Cards クラスの個体であるという記述を RDF に追加する。前項で作成した RDF

に「このリソースは MTG オントロジーの個体である」のトリプルを追加する。

```
<rdf:type rdf:resource="&mtg:Cards"/>
```

その後、Protégé へのインポートを行った。その結果を図 10 に示す。

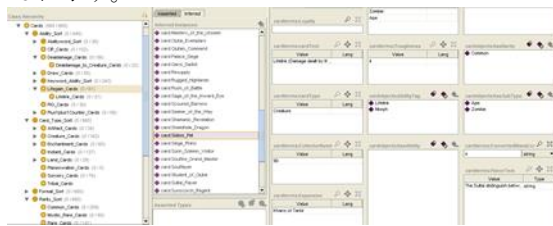


図 10 カード RDF のインポート

インポートしたカードデータを、推論機構を用いて推論した結果を図 11 に示す。

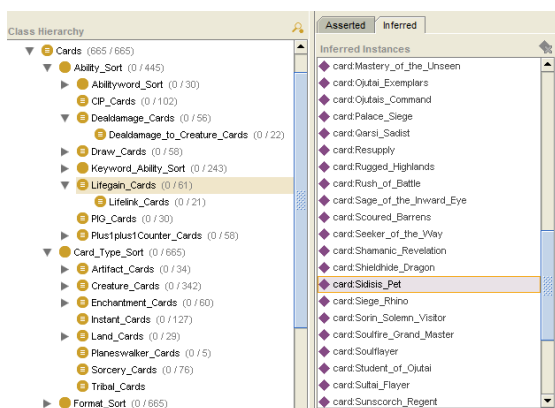


図 11 推論の結果

推論の結果より、「665 種類のカード中にクリーチャー・カードが 342 枚あり、大半を占めている。」といったことや、「ダメージを与えるカードが 56 枚あり、内 22 枚はクリーチャーに対してダメージを与えるカードである。」といった様々なことがわかる。また、「Lifelong」という語は「与えたダメージと同じ点数ライフを回復する。」という意味であるが、テキストベースの検索では「Gain Life」といった語が含まれていなかったため検索結果に表示されなかった。しかし、クラス階層を用いて「Gain Life」の下位クラスとして「Lifelong」を記述することによって「Lifelong」のカードも回復するカードとして推論された。このようにしてオントロジーを作成することによってカードの機械処理を実現した。

6. 考察

(1) RDF の自動生成についての考察

基本的なカードに対してのプログラムによる RDF の自動化は本研究の手法によって達成されたが、一部の MTG 用語が正しくタグ付けされなかった。これはタガーの訓練に使用した MTG コーパスがランダムに選ばれたカードテキストの集合であったため、その MTG 用語が

収録されていなかったためと考えられる。これを解決するには MTG コーパスに加えて、一通りの MTG 用語について記述した辞書データのようなものが必要になると考えられる。

また、一部特殊な形式のカードを RDF 化するには至らなかった。全てのカードを RDF 化するためにはこのような場合に対するデータの形式の選定が必要になる。

(2) オントロジーの推論についての考察

本研究では、オントロジーによる能力に関する分類に AbilityTag を採用し、これによってつけられた能力用のタグの組み合わせによって能力に関する分類をする手法を取った。この方法によって、テキストが短いカードに対しては適切な推論が行えた。しかし、テキストが複数あるようなカードではタグが混合してしまうことが見受けられた。これを解決するには、例えば「Life」と「Gain」が一つの文章中でできた場合「Life Gain」と能力用のタグ付けするように、より具体的なタグ付けを行うことが求められると考えられる。

7. まとめ

カードの RDF 作成の自動化については概ね達成できたと言える。オントロジーによる推論についてはカードの分類を達成し、従来のテキストベースの検索では不便であった部分を改善できた。

また課題としては、カードの分類をするクラス間の関係を推論規則によって記述することでカード間の関係も推論できると考えられるが、本研究では記述することができなかった。この課題をクリアし、サービスとして利用できる形になれば、よりプレイヤーに有益な情報を提供することが可能になると考えられる。

謝辞 本研究を進めるにあたり、ご指導いただいた藤井章博教授に感謝いたします。また、日常の議論の中で多くの示唆を頂いた研究室の皆様にも感謝いたします。

参考文献

- 1) 田中哲郎：「どうぶつしょうぎ」の完全解析。研究報告ゲーム情報学(GI), Vol. 2009-GI-22, No.3, p1-8, 2009
- 2) 古井敬大：相手の抽象化による多人数ポーカーの戦略の決定。東京大学院工学系研究科電気系工学専攻 平成 25 年 修士学位論文, 2013
- 3) Protégé, <http://protege.stanford.edu/>(2015 年 2 月 3 日アクセス)
- 4) Gatherer, <http://gatherer.wizards.com/Pages/Default.aspx>(2015 年 2 月 3 日アクセス)
- 5) Steven Bird, Ewan Klein, Edward Loper : 入門 自然言語処理. 株式会社オライリー・ジャパン. 2010